A comparison of approaches for assessing covariate effects in latent class analysis

Heron, Jon; Croudace, Tim J.; Barker, Edward D.; Tilling, Kate

Published in:
Longitudinal and Life Course Studies

DOI:
10.14301/llcs.v6i4.322

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

A comparison of approaches for assessing covariate effects in latent class analysis

Jon Heron School of Social and Community Medicine, University of Bristol, UK
jon.heron@bristol.ac.uk
Tim J. Croudace School of Nursing and Midwifery, University of Dundee, UK
Edward D. Barker Institute of Psychiatry, King's College London, London, UK
Kate Tilling School of Social and Community Medicine, University of Bristol, UK

(Received October 2014 Revised March 2015) http://dx.doi.org/10.14301/llcs.v6i4.322

Abstract

Mixture modelling is a commonly used technique for describing longitudinal patterns of change, often with the aim of relating the resulting trajectory membership to a set of earlier risk factors. When determining these covariate effects, a three-step approach is often preferred as it is less computationally intensive and also avoids the situation where each new covariate can influence the measurement model, thus subtly changing the outcome under study. Recent simulation work has demonstrated that estimates obtained using three-step models are likely to be biased, particular when classification quality (entropy) is poor. Using both simulated data and empirical data from a large United Kingdom(UK)-based cohort study we contrast the performance of a range of commonly used three-step techniques. Bias in parameter estimates and their precision were determined and compared to new bias-adjusted three-step methods that have recently become available. The bias-adjusted three-step procedures were markedly less biased than the simpler three-step methods. Proportional Maximum Likelihood (ML), with its complex-sampling robust estimation, suffered from negligible bias across a range of values of entropy. Whilst entropy was related to bias for all methods considered, there was evidence that class-separation for each pairwise comparison may also play an important role. Under some circumstances a standard three-step method may provide unbiased covariate effects, however on the basis of these results we would recommend the use of bias-adjusted three-step estimation over these standard methods.

Keywords

ALSPAC, latent class analysis, trajectories, bias, three-step

Introduction

The use of mixture models in epidemiological research has increased markedly in recent years, partly due to developments in statistical software packages such as Mplus (Muthén & Muthén, 2012) and Latent Gold (Vermunt & Magidson, 2013) that have brought these complex, computationally intensive techniques within the grasp of the average applied researcher. Mixture models come in various forms; some designed specifically for longitudinal data e.g. Latent Class Growth Analysis or Growth Mixture Models (Muthén & Muthén, 2000) and others such as standard Latent Class Analysis appropriate in either a longitudinal or cross-sectional setting. All models share one feature, the estimation of an underlying categorical latent variable (hereafter referred to as *X*) which is theorized to be the reason for some or all of the patterns of association observed within the dataset. The procedure will estimate the likely distribution of *X*, namely the number of classes and their prevalence, as well as individual probabilities of

class membership, which describe the allocation of each participant/observation to each latent class under the estimated model. Many stopping rules, e.g. entropy (Ramaswamy, DeSabro, & Robinson, 1993), Bayesian Information Criterion (BIC) (Schwarz, 1978), Bootstrap Likelihood Ratio Test (BLRT) (Nylund, Asparouhov, & Muthén, 2007) have been utilized with the goal of determining an adequate number of classes.

In some cases *X* itself is of little interest, for instance its inclusion may be purely to help with some deviation from normality within the data. However, more often estimating *X* is a key focus as it may represent underlying subpopulations who have different characteristics or who may respond differently to some intervention. The analyst will typically estimate *X* on the basis of a few 'class-indicators', such as repeated measures of enuresis (Croudace, Jarvelin, Wadsworth, & Jones, 2003) or cross-sectional symptoms of psychosis (Shevlin, Murphy, Dorahy, & Adamson, 2007) before offering up *X* for further investigation e.g. to understand which early-life factors distinguish between the classes or what is the long-term prognosis of members of each group. It is during this secondary stage where no firm rules have been established with regard to best practice and a number of analytical approaches have been adopted across the applied literature. Despite the relative ease with which one may determine covariate effects within a "one-step" model where the measurement model for *X* is estimated at the same time as the covariate odds-ratios for class-membership, a number of "three-step" procedures are commonly used.

The term "three-step" (Vermunt, 2010) refers to the sequential stages of firstly estimating the mixture model, secondly exporting the salient features of the model to a different statistical package, before finally analysing some derived indicator of class membership in further analysis, e.g. as the outcome in a multinomial logistic regression model. Popular second-step procedures include assigning each participant to their most likely class (Modal Assignment) or incorporating class-assignment uncertainty either by making multiple draws from each participant assignment probabilities (Pseudo-Class Draws, PCD) or using the probabilities themselves as regression weights (Proportional Assignment). All methods aside from the one-step fall under the banner of three-step methods, even if the second step merely involves exporting the data from step one.

Recent simulation work (Clarke & Muthén, 2009) has demonstrated a number of shortcomings of these three-step methods, including substantial parameter bias and over- precise estimates. However, as described by Clarke & Muthén and also Vermunt, the three-step strategy brings a number of advantages including reduced model complexity as well as avoiding the situation where the form (and potentially interpretation) of *X* may alter depending on the covariates/outcomes included in the model. As is often the case, a single mixture model which defines a sub-division of the study population may give rise to a series of related papers so there is clear benefit to having a consistent, unchanging assignment of the study participants.

In a recent paper, Vermunt (Vermunt, 2010) has brought applied analysts a new alternative by devising a pair of refined three-step procedures. Using standard mixture-modelling output which describes the agreement between the estimated and underlying latent measure, the third step of a three-step procedure can be adjusted to remove the measurement error induced through estimation of the latent measure in step two. Bias and precision are seen to be improved, but crucially the latent class assignment is unchanged, thus a succession of different models can be examined without impacting on the formulation of *X*.

The aim of the current paper is to investigate how these estimation approaches perform in practice, when applied to the analysis of trajectories of conduct problems in childhood (Barker & Maughan, 2009) derived using data from the Avon Longitudinal Study of Parents and Children (ALSPAC), a UK-based birth-cohort. The latent grouping produced in the original manuscript has since been utilized in a number of follow-up publications (Barker, Oliver, & Maughan, 2010; Heron et al., 2013a; Heron et al., 2013b; Kretschmer et al., 2014; Oliver, Barker, Mandy, Skuse, & Maughan, 2011; Stringaris, Lewis, & Maughan, 2014) in which a range of one- and three-step procedures have been employed in order to examine further risk factors for non-normative development or to study late problematic outcomes in those exhibiting different patterns of conduct problem behaviour. In the current manuscript we select a single covariate (gender) in order to

compare results obtained using the range of methods now available. Observations are subsequently verified through simulation.

Methods

Participants

The sample comprised participants from the Avon Longitudinal Study of Parents and Children (ALSPAC) (Boyd et al., 2013; Fraser et al., 2013; Golding, Pembrey, & Jones, 2001). ALSPAC is an ongoing population-based cohort study in the South-West of England. Pregnant women resident in the former Avon Health Authority (which included the city of Bristol), who had an estimated date of delivery between 1 April 1991 and 31 December 1992, were invited to take part, resulting in a cohort of 14,541 pregnancies which resulted in 13,796 singletons and first-born twins who were alive at one year of age. Detailed information about ALSPAC is available online (http://www.bris.ac.uk/alspac) and the study website also contains details of all the data that is available through a fully searchable data dictionary (http://www.bristol.ac.uk/alspac/researchers/data-access/data-dictionary/). Ethical approval for the study was obtained from the ALSPAC Law and Ethics Committee and local Research Ethics Committees.

Outcome - Conduct Problem (CP) trajectories during childhood

The derivation of CP trajectories has been reported previously (Barker & Maughan, 2009). Briefly, Latent Class Growth Analysis was applied to six assessments of mother-reported CP, spanning the age period from four to 13 years, using the 'Conduct Problem' subscale of the Strengths and Difficulties Questionnaire (Goodman, 2001; Goodman & Scott, 1999) The sum-score at each wave was dichotomized at the standard threshold of four or more (Goodman, 2001), yielding six binary indicators. The four resulting trajectories were described as "Low" (72.4%), "Childhood Limited" (CL, 11.8%), "Adolescent Onset" (AO, 7.8%) and "Early-Onset Persistent" (EOP, 8.0%). Proportions quoted are for the complete-case sample (n = 4,659) following modal assignment. Entropy for this model was 0.730.

Exposure

For these models we will focus on offspring sex, which is coded 0 'female', 1 'male' so that parameter estimates indicate the extent to which boys have greater log-odds compared with girls of being in the comparison class.

Statistical methods

Whilst "C" is often used when referring to the latent variable within a latent class model, here we adopt the notation used in Vermunt (2010). We use X to denote the underlying latent variable and W for any predicted classification obtained during the second step of a three-step estimation method. Latent class indicators for subject i are denoted by $\boldsymbol{Y}_i$ and a covariate (predictor of class-membership) by Z_i (i.e. sex in the empirical example).

Empirical models

The effect of sex on latent class variable X (conduct trajectory class) was assessed using a range of one- and three-step methods, each time treating X as a four-category multinomial outcome. Of interest was both the magnitude of the main effects of sex, given by log-odds ratios, and their standard errors. As it is customary to approach these models with the mind-set that these classes are all inherently different in some way, we chose to make comparisons between all classes rather than just deriving parameter estimates with reference to the normative (Low) group. For each comparison we examine percentage deviation from the one-step results, defined to be the difference between each three-step result and those derived from the one-step method, expressed as a percentage of the one-step estimates. We note here that we are making the assumption that the one-step results are correct and for our empirical models we do not know this to be the case.

The following methods were compared:

One-step estimation - The direct effect of sex on X was estimated by incorporating this independent variable into the original mixture model. Estimation was carried out using Mplus version 7.1 (Muthén & Muthén, 2012).

Three-step methods - With all three-step methods the first step entails the estimation of an unconditional mixture model, i.e. a measurement model for latent class X in the absence of any potential covariates. The output from this first step consists of a set of class-assignment probabilities – denoted $P(X = t \mid \boldsymbol{Y}_i)$ – for each respondent. Respondents with the same set of responses for class indicators $\boldsymbol{Y}_i$ are given an identical set of class-assignment probabilities, however depending on the three-step method chosen, such respondents

may not all be assigned to the same class. During step-two these data are used to derive the nominal variable *W,* which is then used as the dependent variable in the final step. Here the methods chosen adopt one of two alternative step-two procedures – Modal Assignment and Proportional Assignment. We first discuss their standard use before describing the bias-adjusted approaches.

Modal Standard - Perhaps the most commonly-used three-step method, the second step entails assigning each respondent to their most likely class (the class for which $P(X = t \mid \mathbf{Y}_i)$ is greatest). In step three this classification *W* becomes the nominal dependent variable in a multinomial logistic regression analysis. Whilst we use Latent Gold for all three-step models described, this model can be estimated in mainstream statistical software such as Stata and SPSS.

Proportional Standard - In contrast to modal assignment, three-step methods based on proportional assignment incorporate the class-assignment probabilities. Proportional Assignment involves stacking ones' class-assignment probabilities so that each respondent has multiple rows of data (one row per class). An additional column is created which indexes these classes. For step-three a multinomial logistic regression model is estimated with this class-index as the dependent variable and the column of assignment probabilities used as regression weights (this method is also known as "Probability Weighting"). This model is also estimable in Stata with the assignment probabilities defined to be "importance weights" and in SPSS through the use of frequency weights.

Modal ML and Proportional ML - The three-step methods Modal Standard and Proportional Standard suffer from two limitations. Firstly they assume a perfect relationship between the classification *W* derived in step two and the unmeasured latent variable *X*, and secondly they fail to account for the fact that *X* is latent so its true values are unknown. Vermunt (2010) devised a pair of bias-adjusted estimation methods, referring to these as "Modal ML" and "Proportional ML". The estimation of these methods requires the appropriate "D-matrix" containing classification probabilities that describe the relationship between *W* and *X*, or put another way, they quantify the measurement error in *W*. Through the use of this classification matrix, a subsequent latent class estimation - well established as a method for dealing with measurement error in categorical variables - is able to reproduce the quantity of interest, namely the effect of covariate Z_i on *X*. As a consequence of the need for a second latent-class analysis, software options for estimating step three are more limited.

Through simulation work, Proportional ML was observed to produce parameter estimates closer to the one-step (true) results, whilst Modal ML gave more accurate standard errors (SE) - SE's for Proportional ML were slightly too large. Vermunt demonstrated how one might estimate these models in Latent Gold, however Modal ML is also estimable in Mplus, and, since version 7.1, has been simplified through use of the "auxiliary" command. See the supplementary material for further details on the derivation of the D-matrix and the estimation of these models in Latent Gold and Mplus. Finally we note that when the D-matrix for either Modal or Proportional Assignment is equal to the identity matrix the Modal Standard or Proportional Standard estimates are reproduced. In other words, as stated above, standard methods make the assumption that there is no measurement error in W.

Modal ML (robust) and Proportional ML (robust) - In a follow-up publication to Vermunt (2010), Bakk and colleagues (Bakk, Oberski, & Vermunt, 2014) revised the estimation methods for both Modal and Proportional ML. By using a complex-sampling robust estimator to allow for within person clustering (in our empirical example the stacked dataset has four rows per respondent) and a Taylor expansion to better allow for the classification-error uncertainty inherent in the third step estimation, improvements on the original bias-adjusted estimates have been demonstrated, particularly for Proportional ML. Modal ML (robust) and Proportional ML (robust) are both available in Latent Gold version 5.0 however neither can be estimated currently in Mplus (version 7.3).

Simulation models

We sought to replicate the findings from the empirical analysis using a simple simulation study. This enabled us to take control aspects of the model such as entropy and class separation, and furthermore ensure that our chosen one-step model was the appropriate one for the data.

Simulation #1: Relationship between bias and entropy

Had we simulated from a model containing a mixture derived from repeated binary indicator variables it would have been difficult to vary entropy/class-separation in a controlled manner. Consequently, the class indicator used here was a single multimodal continuous variate *Y*. Latent class *X* was then to be regressed on a single binary covariate Z_i giving rise to a pair of log-odds ratios describing the Z_i-by-*X* relationship. The Monte Carlo routine in Mplus was used to simulate the necessary data with further details given below.

Defining the relationship between observed class indicator Y and latent class X

Continuous variate *Y* was simulated to be a mixture of three normal distributions of equal size, located at values -1 (class 1), 0 (Class 2) and 2 (Class 3) as illustrated in Supplementary Figure 1. Variances were constrained equal for all three distributions and were increased incrementally from 0.05 to 0.5 in steps of 0.05 yielding ten different simulation models. A (within-class) variance of 0.05 produces a near-perfect value of entropy (~1.0) and very good class separation. As variance is increased, class-separation is reduced initially for the two closer classes (classes 1 and 2) and ultimately all three classes will be poorly separated. Within-class variance was the only aspect of the model to be varied between simulations. 500 replications were produced for each of the ten models with a constant sample size of 5,001. Preliminary work indicated acceptable coverage and bias for the one-step model when using this number of replications.

*Defining the relationship between Covariate Z_i and latent class **X***

The association between binary covariate Z_i and three category nominal outcome *X* can be described as a six-cell contingency table. Consequently, five quantities (in addition to the sample size) are required to fully describe these data. For the set-up used in Mplus, the following details were needed: the proportion of people in the Z_i = 0 group; two log-odds ratios defining the relationship between Z_i and *X*; and two logits to define the class distribution *X* in the unexposed group (Z_i=0). Here we opted for three classes of equal size (n = 1,667). The proportions exposed to Z_i within each class were as follows: class 1 (517/1,667 = 31.0%), class 2 (417/1,667 = 25%), class 3 (317/1,667 = 19%). This results in a covariate Z_i with 25.01% prevalence and log-odds ratios of 0.649 for class 1 and 0.351 for class 2 (with reference to class 3), giving a log-odds ratio of 0.298 for class 1 with reference to class 2. In other words, relative to class 3, exposure to covariate Z_i would convey moderately increased log-odds of being in class 2, and a greatly increased log-odds of being in class 1. Finally, the chosen cell counts imply a class-distribution of *X* of 30.67%/33.33%/36.0% among those unexposed to Z_i, which can be described as two logits: -0.160 and -0.077.

Analysis of simulated data

Each of the one-step and three-step methods were used to estimate the effect of Z_i on *X* for each simulated dataset. This was facilitated through use of the brew package (Horner, 2011) in R (R Core Team, 2014). All parameter estimates were imported into Stata version 13.1 (StataCorp., 2013) where the –simsum– routine (White, 2010) was employed to derive the measure of bias relative to the true regression parameters (0.649, 0.351 and 0.298). We also compared estimate precision by calculating the SD in each parameter estimate across the 500 simulated datasets.

Simulation #2: Relationship between bias and pairwise class separation

Analysts tend to focus on entropy as a single summary measure of class assignment uncertainty for the whole model, however it is often the case that some large classes are well defined with other smaller classes being less so. In this case, it will be the large classes driving entropy, and not all class-comparisons will have the same degree of accuracy. Maitra and Melnykov provide equations (equation 2.1 in Maitra & Melnykov, 2010) for deriving what they refer to as cluster-overlap when estimating a Gaussian mixture model. For each pair of classes, the cluster-overlap is defined as the sum of two misclassification probabilities for the overlap with class *i* when considering class *j*, and vice versa. Hence a pairwise measure of cluster-overlap is readily available and is given by the sum of the [i,j] and [j,i] elements of the "D-matrix". This formally defined measure of cluster-overlap is essentially the opposite of what we have been referring to more loosely as class-separation. For a pair of classes with good separation, overlap will be close to zero. In contrast, independence between *X* and *W* would yield overlap of 2/(# classes), with a more complex *X*-*W* relationship producing potentially greater values, though ultimately bounded by 2.

We sought to investigate the role that cluster-overlap has on the bias of our estimates. Here, we focus on the first comparison (class 1 versus class 3) for which the covariate had the largest effect in the original simulation (log odds = 0.649). For a given value of entropy, the association between parameter bias and pairwise class-overlap is confounded by the magnitude of the covariate effects. Consequently we re-simulated the data after permuting the ordering of the classes. This was done keeping both entropy AND the covariate-effects constant and only works because our three classes were simulated to be of equal size (otherwise the permutation would alter entropy). If we label the original simulation model as "123" reflecting the ordering of the classes at locations -1, 0 and 2, then permuting the classes to orders "312" and subsequently "231" enables us to vary class-separation as shown in figure 3. Note that there are three other possible class orderings, "132", "213" and "321", which produce the same three measures of cluster-overlap and the same levels for bias ("123" is equivalent to "321" etc.). Following the simulation of these new data, the same analytical steps were performed as for Simulation #1. Parameter estimate bias was calculated and its relationship with cluster-overlap was examined.

Results

Empirical example

Estimated sex effects for each pair of latent classes are shown in table 1. Figures in parentheses show percentage deviation from the one-step results. As the entropy for the original mixture model was not particularly high (0.730), previous simulation work would predict that standard three-step methods would be inaccurate, typically under-estimating the effects of sex and also being overly-precise since these methods do not capture the uncertainly in estimated class assignment.

Parameter estimates

For all class comparisons, the standard three-step methods produce estimates closer to the null than the one-step results. Estimates obtained using Proportional ML are consistently within 1 or 2% of the one-step results. Modal ML estimates are more variable, and are substantially higher than the one-step for the comparison of classes Childhood Limited and Early Onset Persistent. Unsurprisingly, the use of robust SE's has no effect here.

Standard errors

Again, as expected, the standard three-step methods are overly precise with SE's up to 32% and 58% lower that the one-step for Modal and Proportional Standard respectively. Proportional ML severely over-estimates SE, however the new complex-sampling robust variance estimator demonstrates a marked improvement here. The robust estimator has little effect on Modal ML, with all SE's being moderately raised compared to one-step and Proportional ML (robust).

Summary of empirical findings

The three-step methods chosen produced a wide range of estimates for the parameters and their standard errors. What is apparent is that deviations relative to the one-step values are typically lower, particularly for the standard errors, when comparing pairs of classes which have better separation. Like many longitudinal mixture models, the analysis of conduct problems produced patterns of trajectories which have been described previously as a soldier's bed or cat's cradle (Sher, Jackson, & Steinley, 2011) in other words high and low relatively flat trajectories and a pair of trajectories which cross midway through the time period. Here the classes which cross (AO and CL) are less well separated, whilst the two persistent classes (Low and EOP) have little overlap. This appears to be reflected in the consistency of their estimates across the methods.

Simulation #1: Relationship between bias and entropy

Unconditional three-class mixture models estimated on each simulated dataset reported the following entropy values (averaged across 500 datasets): 0.98, 0.91, 0.85, 0.79, 0.75, 0.70, 0.67, 0.63, 0.61 and 0.58. Figure 1 shows the relationship between entropy and the percentage bias obtained in the parameter estimates and figure 2 shows estimated precision (SD of estimate across datasets) for each method.

When comparing results from bias-adjusted methods our findings were consistent with recent simulation work (Bakk et al., 2014). Modal ML and Modal ML (robust) results were almost identical in both bias and precision, likely due to the large sample size in our examples. In contrast (as expected), there was a marked increase in precision with Proportional ML (robust). Standard errors for Proportional ML (robust) were within 3% of the

one-step values for all values of entropy whereas for non-robust Proportional ML the standard errors were in one instance 86% higher than those obtained using a one-step approach. On the basis of these results we would caution against the use of Proportional ML without robust standard errors. Here we report results only for the two more recent methods – Modal ML (robust) and Proportional ML (robust) – however a full set of results are available on request. To facilitate a clearer comparison of these two methods, we have reproduced the figures after removing the standard methods to enable the y-axis to be restricted (see supplementary material).

Parameter estimate bias

Due to the location of the three classes, reduction in entropy initially impacts on the comparison of class 1 versus class 2 (third comparison) followed by the other two comparisons. We observe both positive and negative bias in this example, however we note that estimates affected by positive bias will be bounded by the maximum value of the true log-odds ratios – in this case 0.649 (Bolck, Croon, & Hagenaars, 2004). The standard three-step methods are badly affected by the reducing entropy, with Modal Standard fairing slightly better but still producing unacceptable levels of bias unless entropy is very high. Both bias-adjusted three-step methods produce estimates with a low level of bias for all three class comparisons and across the wide range of entropy values considered.

We see that for the second comparison the bias for standard three-step methods appears to decrease for lower values of entropy. This phenomenon is merely an artefact of our chosen simulation. As entropy reduces, the distinction between classes 1 and 2 is the first to become affected such that class 1 becomes more similar to class 2 and vice versa. Since class 1 is more strongly associated with the covariate, our second comparison (class 2 versus class 3) is boosted, partially offsetting the negative-bias present in both standard methods.

Standard Errors

Decreasing entropy should increase uncertainty and accordingly we observe a reduction in precision for the (correct) one-step model. Standard errors for Proportional ML (robust) closely match the one-step values with Modal ML (robust) giving slightly higher values. What is most apparent from these figures is that the standard three-step approaches are failing to capture the increasing uncertainty, in fact in this example Proportional Standard becomes more precise as the level of assignment uncertainty increases.

Simulation #2: Relationship between bias and pairwise class separation

Table 2 shows the resulting biases for this second set of simulations. Output is restricted here to the five highest values of entropy – typically the range in which an analyst might be considering the use of a standard three-step method. These results are split into two since methods using Modal and Proportional assignment will have a different D-matrix and hence a different value for class-separation for the same dataset. We see that for very high levels of entropy (>> 0.9) there is little detriment to using any modelling approach. However unacceptable (>10%) levels of bias in the parameter estimate is present when entropy is still extremely high (0.91) if the class overlap is moderate, and in contrast, *less* bias for *lower* entropy (0.75 – 0.80) when a particular pair of classes has a good degree of separation. Whilst these results are limited in scope, they suggest that a decision based solely on entropy may be unwise.

Discussion

Using an empirical example from a large UK birth cohort and a limited set of simulations we have compared the estimate effect of a single covariate on latent class membership using various three-step approaches commonly used in applied papers from the fields of psychology, epidemiology and medicine. Our findings are consistent with previous simulations showing that standard three-step methods can produce results which are both biased and overly precise, particularly when entropy is poor. What this study adds is the suggestion that entropy, a single-summary measure of classification quality, is only part of the story and we would advise caution regarding a modelling strategy based solely on its value, for instance whether it exceeds an arbitrary threshold such as 0.8 or 0.9.

We have demonstrated that for extremely high values of entropy it remains possible for individual class comparisons to be biased if the separation between those classes is poor. In contrast, when entropy is low, some class comparisons may be unbiased if their separation is good relative to the

rest of the model. When faced with the worst-case scenario of a combination of low entropy and poorly separated classes, only proportional ML (robust), of the three-step methods, appears to fare well, however previous simulations suggest that for extremely low entropy all three-step methods may be flawed (Bakk, Tekle, & Vermunt, 2013; Vermunt, 2010) leaving the one-step method as the only option for obtaining unbiased estimates. Our simulation focussed on what would be regarded as a large sample size for this type of analysis and this is likely to be an explanation for the strong performance of proportional ML (robust) across the whole range of entropy considered.

It is clear from our results that pairwise class-separation may play an important role in determining the level of bias in the standard three-step methods, although we are unable to make recommendations with regard to acceptable thresholds. There is a strong link between separation and entropy, and separation will be also affected by the number of classes present and their relative positioning. Thus, derivation of thresholds for class-separation will be challenging. In our view further efforts would be better directed at facilitating the use of bias-adjusted three-step methods within mainstream statistical software.

In our empirical example we focussed on the respondents with a full set of class indicators. Whilst we observed good agreement between the one-step and the robust ML three-step methods our sample used for analysis consists of merely one third of ALSPAC hence our estimates may not generalise to the broader sample of those who enrolled. Here we make a number of observations in relation to this since the topic of missing data in the context of three-step estimation is currently unexplored.

Firstly, Full Information Maximum Likelihood (FIML) permits the inclusion of partial respondents based on the missing-at-random (MAR) assumption. However, as entropy for such a model would be expected to be lower due to additional uncertainty surrounding these incomplete observations, there is the potential for this to offset gains made through the use of a larger, more representative sample. Alternative approaches include focussing on a sample for which a rich set of class-indicators are available and using a weighting method, e.g. Inverse Probability Weighting (IPW), to adjust for any potential selection bias. IPW has recently been shown to be a useful technique when used in combination with other missing data methods (Seaman, White, Copas, & Li, 2012). Secondly, when using likelihood-based methods to deal with missing data, one may condition on predictors of missingness to strengthen the MAR assumption. Were covariate Z_i to be an important predictor of dropout as well as being an exposure of interest, one would surmise that only the one-step method would achieve an unbiased result. Finally, FIML-based mixture modelling can only deal with missing covariate information (incomplete Z) in a rather simple setting and by making potentially undesirable distributional assumptions. A clear advantage of the treat-as-observed approach of Modal Standard is the ease with which one may then incorporate classification W into a multiple imputation model where any covariate missingness can be dealt with. Future developments could focus on a toolkit for the applied researcher that allows bias-adjusted estimation of the Z_i-by-X association with a range of currently state-of-the-art missing data treatments.

Acknowledgements and funding

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and Jon Heron and Kate Tilling will serve as guarantors for the contents of this paper. This research was specifically supported by the Medical Research Council [grant number G1000726]. Kate Tilling works in a unit that receives funding from the UK Medical Research Council and the University of Bristol (MC_UU_12013/5).

References

Bakk, Z., Oberski, D. L., & Vermunt, J. K. (2014). Relating Latent Class Assignments to External Variables: Standard Errors for Correct Inference. *Political Analysis, 22*, 520-540. http://dx.doi.org/10.1093/pan/mpu003

Bakk, Z., Tekle, F. T., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology, 43*, 272-311. http://dx.doi.org/10.1177/0081175012470644

Barker, E. D., & Maughan, B. (2009). Differentiating early-onset persistent versus childhood-limited conduct problem youth. *American Journal of Psychiatry, 166*, 900-908. http://dx.doi.org/10.1176/appi.ajp.2009.08121770

Barker, E. D., Oliver, B. R., & Maughan, B. (2010). Co-occurring problems of early onset persistent, childhood limited, and adolescent onset conduct problem youth. *Journal of Child Psychology and Psychiatry, 51*, 1217-1226. http://dx.doi.org/10.1111/j.1469-7610.2010.02240.x

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating Latent Structure Models with Categorical Variables: One-Step Versus Three-Step Estimators. *Political Analysis, 12*, 3-27. http://dx.doi.org/10.1093/pan/mph001

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S., & Davey Smith, G. (2013). Cohort Profile: The 'Children of the 90s'-the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology, 42*, 111-127. http://dx.doi.org/10.1093/ije/dys064

Clarke, S. L., & Muthén, B. (2009). Relating Latent Class Analysis Results to Variables Not Included in the Analysis. Retrieved from www.statmodel2.com/download/relatinglca.pdf

Croudace, T. J., Jarvelin, M. R., Wadsworth, M. E., & Jones, P. B. (2003). Developmental typology of trajectories to nighttime bladder control: epidemiologic application of longitudinal latent class analysis. *American Journal of Epidemiology, 157*, 834-842. http://dx.doi.org/10.1093/aje/kwg049

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A., Ring, S., Nelson, S. M., & Lawlor, D. A. (2013). Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology, 42*, 97-110. http://dx.doi.org/10.1093/ije/dys066

Golding, J., Pembrey, M., & Jones, R. (2001). ALSPAC--the Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatric and Perinatal Epidemiology, 15*, 74-87. http://dx.doi.org/10.1046/j.1365-3016.2001.00325.x

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*, 1337-1345. http://dx.doi.org/10.1097/00004583-200111000-00015

Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? *Journal of Abnormal Child Psychology, 27*, 17-24. http://dx.doi.org/10.1023/A:1022658222914

Heron, J., Barker, E. D., Joinson, C., Lewis, G., Hickman, M., Munafo, M., & Macleod, J. (2013a). Childhood conduct disorder trajectories, prior risk factors and cannabis use at age 16: birth cohort study. *Addiction, 108*, 2129-2138. http://dx.doi.org/10.1111/add.12268

Heron, J., Maughan, B., Dick, D. M., Kendler, K. S., Lewis, G., Macleod, J., Munafo, M., & Hickman, M. (2013b). Conduct problem trajectories and alcohol use and misuse in mid to late adolescence. *Drug and Alcohol Dependence, 133*, 100-107. http://dx.doi.org/10.1016/j.drugalcdep.2013.05.025

Horner, J. (2011). *Templating Framework for Report Generation. R package version 1.0-6.* . Retrieved from http://CRAN.R-project.org/package=brew

Kretschmer, T., Hickman, M., Doerner, R., Emond, A., Lewis, G., Macleod, J., Maughan, B., Munafo, M. R., & Heron, J. (2014). Outcomes of childhood conduct problem trajectories in early adulthood: findings from the ALSPAC study. *European Child & Adolescent Psychiatry, 23*, 539-549. http://dx.doi.org/10.1111/add.12268

Maitra, R., & Melnykov, V. (2010). Simulating Data to Study Performance of Finite Mixture Modeling and Clustering Algorithms. *Journal of Computational and Graphical Statistics, 19*, 354-376. http://dx.doi.org/10.1198/jcgs.2009.08054

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism, clinical and experimental research, 24*, 882-891. http://dx.doi.org/10.1111/j.1530-0277.2000.tb02070.x

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide. 7th Edition*: Los Angeles, California: Muthén & Muthén.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modelling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 535-569. http://dx.doi.org/10.1080/10705510701575396

Oliver, B. R., Barker, E. D., Mandy, W. P., Skuse, D. H., & Maughan, B. (2011). Social cognition and conduct problems: a developmental approach. *Journal of the American Academy of Child and Adolescent Psychiatry, 50*, 385-394. http://dx.doi.org/10.1016/j.jaac.2011.01.006

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org

Ramaswamy, V., DeSabro, W., & Robinson, W. (1993). An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data. *Marketing Science, 12*, 103-124. http://dx.doi.org/10.1287/mksc.12.1.103

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics, 6*, 461-464. http://dx.doi.org/10.1214/aos/1176344136

Seaman, S. R., White, I. R., Copas, A. J., & Li, L. (2012). Combining multiple imputation and inverse-probability weighting. *Biometrics, 68*, 129-137. http://dx.doi.org/10.1111/j.1541-0420.2011.01666.x

Sher, K. J., Jackson, K. M., & Steinley, D. (2011). Alcohol use trajectories and the ubiquitous cat's cradle: cause for concern? *Journal of Abnormal Psychology, 120*, 322-335. http://dx.doi.org/10.1037/a0021813

Shevlin, M., Murphy, J., Dorahy, M. J., & Adamson, G. (2007). The distribution of positive psychosis-like symptoms in the population: a latent class analysis of the National Comorbidity Survey. *Schizophrenia research, 89*, 101-109. http://dx.doi.org/10.1016/j.schres.2006.09.014

StataCorp. (2013). *Stata Statistical Software: Release 13.*: College Station, Texas: StataCorp LP.

Stringaris, A., Lewis, G., & Maughan, B. (2014). Developmental pathways from childhood conduct problems to early adult depression: findings from the ALSPAC cohort. *British Journal of Psychiatry, 205*, 17-23. http://dx.doi.org/10.1192/bjp.bp.113.134221

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis, 18*, 450-469.

Vermunt, J. K., & Magidson, J. (2013). *Latent GOLD 5.0 Upgrade Manual*: Belmont, Massachusetts: Statistical Innovations Inc.

White, I. R. (2010). simsum: Analyses of simulation studies including Monte Carlo error. *The Stata Journal, 10*, 369-385.

Table 1. Parameter estimates for the effect of gender on the four-class multinomial outcome describing trajectories of conduct problems through childhood.

			Methods based on modal assignment			Methods based on proportional assignment		
Reference class	Comparison class	one-step	Modal standard	Modal ML	Modal ML (robust)	Prop standard	Prop ML	Prop ML (robust)
Parameter estimates for effect of sex								
Low	CL	0.388	0.290 (-25.3)	0.407 (4.9)	0.407 (4.9)	0.197 (-49.2)	0.383 (-1.3)	0.383 (-1.3)
Low	AO	-0.125	-0.062 (-50.6)	-0.158 (26.4)	-0.158 (26.4)	0.019 (-115.0)	-0.127 (1.6)	-0.127 (1.6)
Low	EOP	0.303	0.220 (-27.5)	0.279 (-7.9)	0.278 (-8.3)	0.232 (-23.6)	0.301 (-0.7)	0.301 (-0.7)
CL	EOP	-0.084	-0.070 (-16.4)	-0.128 (52.4)	-0.129 (53.6)	0.034 (-141.0)	-0.083 (-1.2)	-0.083 (-1.2)
AO	EOP	0.429	0.281 (-34.4)	0.437 (1.9)	0.436 (1.6)	0.213 (-50.4)	0.427 (-0.5)	0.428 (-0.2)
AO	CL	0.513	0.352 (-31.5)	0.566 (10.3)	0.565 (10.1)	0.178 (-65.2)	0.510 (-0.6)	0.510 (-0.6)
Standard error for above parameter estimate								
Low	CL	0.125	0.093 (-25.9)	0.132 (5.6)	0.132 (5.6)	0.085 (-32.0)	0.176 (40.8)	0.121 (-3.2)
Low	AO	0.151	0.111 (-26.7)	0.169 (11.9)	0.168 (11.3)	0.099 (-34.6)	0.236 (56.3)	0.151 (0.0)
Low	EOP	0.127	0.109 (-13.9)	0.130 (2.4)	0.130 (2.4)	0.109 (-14.3)	0.145 (14.2)	0.124 (-2.4)
CL	EOP	0.171	0.135 (-21.2)	0.179 (4.7)	0.179 (4.7)	0.128 (-25.0)	0.219 (28.1)	0.166 (-2.9)
AO	EOP	0.203	0.148 (-27.2)	0.225 (10.8)	0.225 (10.8)	0.138 (-32.0)	0.305 (50.2)	0.201 (-1.0)
AO	CL	0.200	0.136 (-32.1)	0.222 (11.0)	0.221 (10.5)	0.085 (-57.6)	0.328 (64.0)	0.199 (-0.5)

Figures in brackets indicate percentage deviation from the one-step results
CL: Childhood Limited, AO: Adolescent Onset, EOP: Early Onset Persistent

Table 2. The relationship between bias and class-separation for the simple and bias-adjusted three-step methods (effect of covariate Z on class 1 relative to class 3)

		Methods based on modal assignment					Methods based on proportional assignment				
Entropy	Class order	Class overlap	Modal standard		Modal ML (robust)		Class overlap	Proportional standard		Proportional ML (robust)	
			Estimate	% bias	Estimate	% bias		Estimate	% bias	Estimate	% bias
0.979	123	0.00	0.642	-1.1%	0.646	-0.6%	0.00	0.640	-1.4%	0.646	-0.6%
	231	0.00	0.639	-1.6%	0.644	-0.8%	0.00	0.637	-2.0%	0.644	-0.8%
	312	0.03	0.628	-3.2%	0.645	-0.7%	0.04	0.620	-4.5%	0.645	-0.6%
0.912	123	0.00	0.630	-3.1%	0.646	-0.6%	0.00	0.622	-4.3%	0.646	-0.6%
	231	0.00	0.620	-4.6%	0.643	-1.0%	0.00	0.609	-6.2%	0.644	-0.9%
	312	0.11	0.571	-12.1%	0.646	-0.5%	0.17	0.535	-17.7%	0.646	-0.5%
0.849	123	0.00	0.615	-5.2%	0.645	-0.7%	0.00	0.602	-7.2%	0.645	-0.6%
	231	0.01	0.598	-7.9%	0.642	-1.1%	0.01	0.579	-10.8%	0.644	-0.9%
	312	0.20	0.516	-20.5%	0.648	-0.3%	0.29	0.457	-29.7%	0.647	-0.4%
0.795	123	0.00	0.603	-7.2%	0.645	-0.7%	0.00	0.585	-10.0%	0.645	-0.7%
	231	0.03	0.576	-11.2%	0.643	-0.9%	0.04	0.547	-15.7%	0.644	-0.8%
	312	0.26	0.469	-27.8%	0.646	-0.6%	0.38	0.394	-39.3%	0.648	-0.3%
0.748	123	0.00	0.592	-8.9%	0.645	-0.7%	0.00	0.568	-12.5%	0.645	-0.7%
	231	0.05	0.554	-14.7%	0.644	-0.8%	0.07	0.515	-20.6%	0.645	-0.7%
	312	0.32	0.430	-33.7%	0.644	-0.8%	0.45	0.344	-47.0%	0.648	-0.2%

Estimate = average point estimate across 500 replications. % bias = percentage bias relative to true value of 0.649. i.e. (100%*estimate – true-value)/true-value)

Figure 1. Estimated parameter percentage bias = 100%*((estimate – true-value)/ true-value)

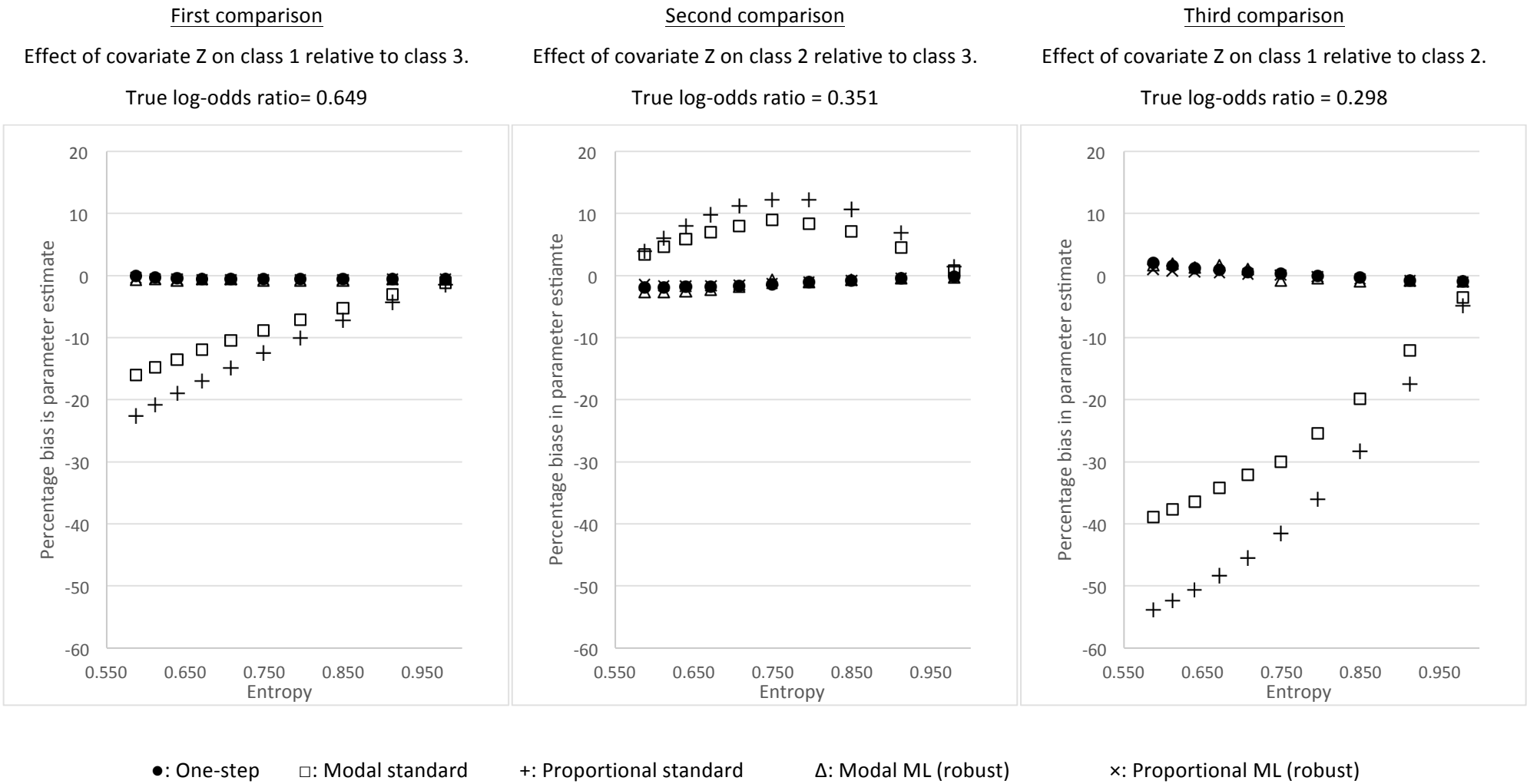

Figure 2. Estimated empirical SE (Standard Deviation of the point estimates across 500 replications) for each method

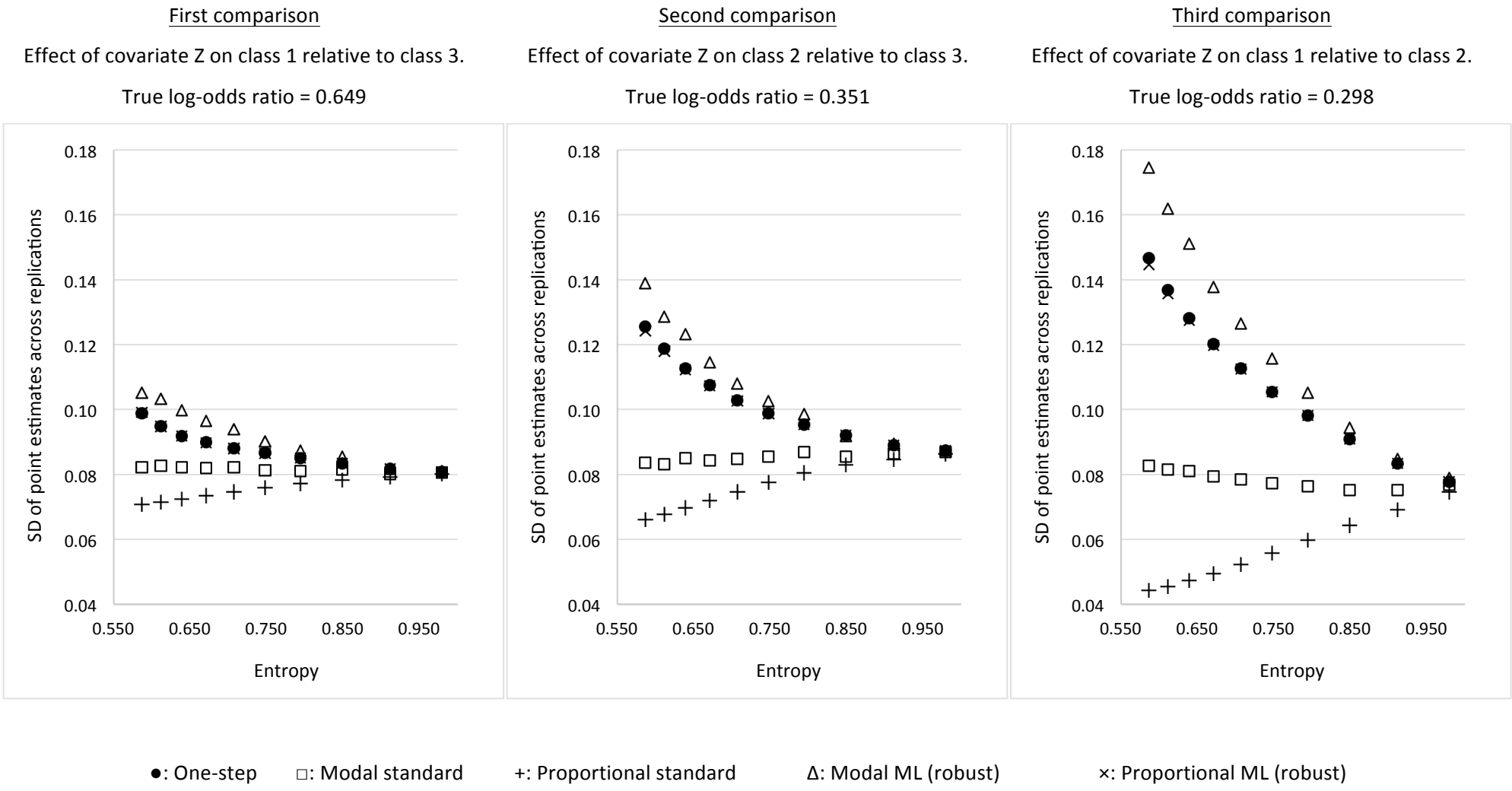

Figure 3. Permutation of the class ordering to control class separation